Supplement to Kostadinov et al. **" Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and earth system models "**

Part 1. Details of the DFT Analysis

The per-pixel time series of 1-degree PFT data (Sect. 2.2 and 2.3) is a 60 x 1 vector $\boldsymbol{x}$, consisting of monthly sample points, x_t; t = 1,2,3,…N, N = 60. For the purposes of Fourier analysis, we consider one year a unit of time, thus the sampling frequency is f_s = 12 times per year. The complex-valued discrete Fourier transform (DFT) of $\boldsymbol{x}$, denoted $\boldsymbol{y}$, was obtained using the Fast Fourier Transform (FFT) algorithm as follows (*Heinzel et al., 2002*):

$$y_m = \sum_{t=0}^{N-1} x_t e^{\frac{-2\pi i * m * t}{N}}, m = 0,1,2,...N-1 \quad \text{(S1)}$$

No windows were applied (i.e. a top-hat window was implicitly applied). The two-sided DFT transform vector $\boldsymbol{y}$ is composed of the y_m elements. The frequency resolution of the FFT is thus f_s/N = 0.2 yr^{-1}, and the frequencies to which the elements of **y** correspond are f = 0, 0.2, 0.4, 0.6, … yr^{-1} up to the highest resolvable frequency – the Nyquist frequency $f_{Nyquist} = f_s/2$ = 6 yr^{-1}. The first element of the $\boldsymbol{y}$ vector is real and equal to the arithmetic mean of the data set $\boldsymbol{x}$; thus it is equal to zero because the mean was subtracted before Eq. S1 was applied. Because of a property of the DFT of real valued input, the first half of vector $\boldsymbol{y}$ contains the same information as the second half and they are complex conjugates of each other. For example if the 2nd element is a + bi, then the last element is a – bi. Because of this conjugate symmetry, only the first half of the vector **y** has to be considered. The power in each frequency is spread to both complex conjugates, so we need to multiply the square of the modulus of each element of the first half of **y** by 2, with the important exceptions of the 1st element (the mean), and in the case of even N - the (N/2 + 1)-*th* element corresponding to the Nyquist frequency (*Heinzel et al., 2002*).

The DFT vector is thus normalized as follows, in order to obtain the power spectrum density, P_{xx} (*Heinzel et al., 2002*) (symbology as above):

$$P_{xx}(f_m) = \left(\frac{k|y_m|^2}{Nf_s} \right); m = 0,1,2,...\frac{N}{2} \quad \text{(S2)}$$

where k = 1 for m = 0 (the signal mean, 0th frequency) and m = N/2 (the Nyquist frequency), and k = 2 for all other m values. P_{xx} has the units of squared data units per cycle per year, that is, power per unit frequency.

The power spectrum, *PS*, can be obtained by multiplying Eq. S2 by the frequency step, f_s/N, i.e. (*Heinzel et al., 2002*):

$$PS(f_m) = \frac{k|y_m|^2}{N^2}; m = 0,1,2,...\frac{N}{2} \quad \text{(S3)}$$

In the above Eq. S3, k takes values as for Eq. S2. The power spectrum has units of power, i.e. squared units of the input data $\boldsymbol{x}$. It represents the power in each frequency band, but not normalized to the frequency bin. Integrating P_{xx} over all frequencies or, equivalently, summing the elements of *PS* results in the variance of $\boldsymbol{x}$. This relationship is known as Parseval's theorem and was verified by testing that the variance computed in the frequency domain matches the time domain variance (normalized by N) to within machine precision. This also verifies that the scaling of the DFT was correct (*Heinzel et al., 2002*). Previous use of Fourier analysis for phenology analysis is documented, for example, in *Moody and Johnson (2001)*. *De Beurs and Henebry (2010)* provide an overview of various phenology analysis methodologies and their advantages and disadvantages.

The annual seasonal cycle was modeled using the fundamental frequency $f = 1$ yr^{-1} and all its available harmonics, i.e. all frequency bands that are its integer multiples, up to the Nyquist frequency of 6 yr^{-1} ($f = 1,2,\ldots 6$ yr^{-1}). That is, the signal was modeled as the summation of sinusoids whose amplitudes and phases are determined by the Fourier coefficients y_m at the respective harmonic frequencies (*MathWorks, 2015*):

$$\hat{\mathbf{x}} = a_0 + a_n \cos(2\pi ft) - b_n \sin(2\pi ft); f = [1;6], f \in \mathbf{Z} \quad \text{(S4)}$$

In the above, $\hat{\mathbf{x}}$ represents the modeled signal, t represents time in years, a_0 corresponds to the signal mean and is equal to y_0 (Eq. S1). The a's are the real part, and the b's – the imaginary part – of the Fourier transform y_m at the corresponding harmonic frequencies, divided by N. Because of conjugate symmetry, the a and b coefficients were doubled for all frequencies except $f = 0$ yr^{-1} and $f_{Nyquist}$. Equation S4 above was applied at a fine temporal sampling of $\Delta t = 1/10^{th}$ of a day approximately. However, the time of maximum is aggregated to a month because the underlining data sets have monthly resolution.

Peak analysis is then applied to $\hat{\mathbf{x}}$ as described in Sect. 2.3, in order to determine the relevant phenological parameters. The percent variance explained by the modeled signal was calculated as the summation of the power spectrum (*PS*, Eq. S3) terms corresponding to $f = 1, 2, 3, 4, 5$, and 6 yr^{-1}, divided the total variance of the input data vector $\boldsymbol{x}$ (Eq. S1). An example of the DFT-based seasonal cycle modeling and peak analysis is illustrated in Fig. S1.

As a verification of the DFT and peak analysis methodology used to derive phenological parameters, we applied the methodology to monthly SeaWiFS photosynthetically available radiation (PAR, mol photons m^{-2} day^{-1}). We used PAR data because it is expected to exhibit strong predictable seasonality in most of the world. For brevity, PAR analysis results are not plotted here, but results are summarized as follows. The derived seasonal amplitude was low near the Equator and higher poleward, as expected. The percent variance explained by the DFT-modeled signal indicated that in most of the world PAR seasonality is well captured by a sinusoidal model and explains most of the variability of the signal. The month of maximum for the primary peak in monthly PAR is December or January in most of the Southern Hemisphere and June, July or August in most of the North Hemisphere. Equatorial locations are expected to have two peaks each year near the equinoxes, but the peaks need not be the same amplitude, so the primary peak occurs in March in some places and August or September - in others. The PAR primary peak duration tends to be highest near to, but not quite at, the Equator (where there are double peaks) and gets progressively shorter at higher latitudes in both hemispheres.

Since missing data (especially at high latitudes) can affect phenological analysis (*Sapiano et al., 2012; Cole et al., 2012*), we investigated the effect of data gaps on the DFT analysis used here. We computed monthly averages of top of the atmosphere (TOA) total solar irradiance (insolation) using the model of *Kostadinov and Gilb (2014)*. Since monthly TOA insolation (and thus PAR) can be exactly 0 W m^{-2} for some months near the solstices during the corresponding polar nights, the same phenological analysis (Sect. 2.3 and this Supplement section) applied to the PFT data sets was also applied to two variants of the monthly TOA data set – the original data set with exact zeros near the Poles, and with those zeros replaced by missing data and interpolated over (as done with gaps in the PFT data sets). When interpolation was used and the zeros were ignored, the signal mean was increased and the amplitude was decreased by small amounts for latitudes above 70^{o}. These effects are small for TOA insolation but could be more significant and could occur at lower latitudes for PAR and especially for PFT and Chl data. Interpolation over missing data can also introduce artificial non-sinusoidal waveforms, which will cause artificial peaks in the DFT spectra and the modeled signal. Observed small decreases of PAR amplitude poleward of ~45^{o} (not shown) could therefore be due to real geophysical phenomena (such as cloudiness patterns), but they may also be modeling artifacts (missing PAR data starts to occur for at least one month per year at these latitudes). Thus, results of analysis presented here at high latitudes or in other areas of frequent missing data must be interpreted with caution (see also Fig. S6).

Part 2. IPCC Models (CMIP5) Data and Phenological Parameters

In addition to analyzing phenology from various ocean color PFT algorithms (Table 1), we also investigated the same phenological parameters in a group of Earth System simulations from the recent Coupled Model Inter-comparison Project CMIP5 (*Taylor et al., 2012*). CMIP5 model output was downloaded from http://pcmdi9.llnl.gov/esgf-web-fe/. We derived phenological parameters from the same five years of "present" historical output (2003 to 2007) of the variable 'phydiat' ("mole concentration of diatoms expressed as carbon in seawater"). Diatom carbon concentration was chosen because it is most similar to the large phytoplankton variables of the satellite algorithms (Table 1); hence only those models that provide this variable are used. The "present" output for 2003-2005 is based on the historical scenario (years 1850 to 2005) forced by observed atmospheric changes (both anthropogenic and natural). The last two years (2006 and 2007) of the "present" output are based on the RCP8.5 scenario (*Riahi et al., 2011*). Table S1 provides details and references for the models. CMIP5 model details and global analysis of ecology for the present and for the 21^{st} century for this same subset of models was presented in *Cabré et al. (2015)*. Molar concentration provided by the models (mol diatom C m^{-3}) was converted to mass concentration (mg C m^{-3}) using the atomic weight of carbon (12.011 g/mol, *Wieser et al., 2013*). All model output was resampled to a 1^{o} grid before applying the DFT calculations (Sect. 2.3). Before computing phenological parameters, biomass values below 0 were set to missing data.

Table S1. Summary of the CMIP5 models that were used in the phenology inter-comparison study presented here. These models were selected because they explicitly model diatom biomass.

Model	*Nutrients*	*Ecology module*	*Phytoplankton variables*	*References*
CESM1-BGC	P, N,Fe,Si	MET	Diatoms, small phytoplankton, diazotrophs	Moore et al. (2004), Moore et al. (2006)
GFDL-ESM2G (M)	P,N,Fe,Si	TOPAZ2	Large phytoplankton (diatoms, greens, and other large eukaryotes), small phytoplankton (prokaryotic picoplankton and nanoplankton), and diazotrophs	Dunne et al. (2013)
HadGEM2-ES	N,Fe,Si	Diat-HadOCC (NPZD)	Diatoms, non-diatoms	Palmer and Totterdell (2001)
IPSL-CM5A-MR	P,N,Fe,Si	PISCES (from HAMOCC5)	Diatoms, nanophytoplankton (non-diatom).	Aumont and Bopp (2006), Séférian et al. (2013)
GISS-E2-H-CC (GISS-E2-R-CC)	N, Fe, Si	NOBM	Diatoms, chlorophytes, cyanobacteria, coccolitophores	Gregg (2008)

Part 3. Statistics of Circular Quantities

Quantities such as day or month of maximum are circular and can be transformed to an angular representation, i.e. angles close to 0 are also close to 2π, and December is temporally close to January. Thus, conventional descriptive statistics are not applicable to such quantities, e.g. for the calculation of the mean and variance of a set of months or angles (*Berens, 2009*). In particular, differences between two months cannot be simply calculated as it is done for linear differences. To resolve this issue and provide meaningful difference and mean and variance estimates of circular quantities used here, we employed circular statistics principles. The signed difference between two months was computed by first representing months in angular form (each month spans 30 degrees, mid-January is at 15^{o}, mid-February at 45^{o}, etc.). These angles were treated as vectors on the units circle and the absolute value of the angle between them was calculated using their dot product. This angle was converted to time units and represented the time difference between two months. The sign of the difference was determined by the sign of the sine of the angle resulting from subtracting the angular representations of the two months. This sign indicates which value leads and which lags in time. By convention, the difference is positive when the first element of the subtraction leads in time.

To calculate the mean month of maximum among the algorithms or models, the months were also represented in angular form and the ***i*** and ***j*** components of the corresponding vectors were averaged separately. The resultant vector's orientation, calculated with the arctangent function, determined the mean months of maximum. This method is consistent with *Berens (2009)*. Variance was computed as one minus the length of this resultant vector (*Allen and Johnson, 1991; Berens, 2009*). Unlike the variance of linear quantities, circular variance is bound between 0 and 1.

Part 4. Methodoloy for Regionally Binned Analyses (North Atlantic)

In addition to the per-pixel analysis of phenology, satellite algorithm and model data were spatially binned and regional analysis was performed for two example *Longhurst (1998)* marine biogeographical provinces (obtained from the Flanders Marine Institute (*Claus et al.*, 2013)), representing regimes of 1) North Atlantic spring bloom region (Westerlies - North Atlantic Drift – NADR, $3.512*10^6$ km^2), and 2) a subtropical gyre region (North Atlantic Subtropical Gyre West - NASW, $5.809*10^6$ km^2). Data for the spatial averages were extracted from the 9-km original imagery for the satellite data (PhytoDOAS data were first resampled to 9 km using nearest neighbor interpolation) and from the 1-degree data for the CMIP5 models. The scalar regional value was obtained by calculating the weighted pixel average, the weights being pixel area. Chl-based data (OC4v6 Chl and PhytoDOAS, Table 1) were averaged in log space. If zeros were present within the region, those pixels were excluded from the log-space averaging, but the final spatial average was weighted by the number of non-zero pixels divided by the number of valid pixels (*Habib, 2012*). Monthly climatological time series were also computed for the Longhurst provinces by averaging the data for all years of a given month (2003 to 2007) after the spatial binning. The same DFT-based phenological analysis was performed on the regionally binned complete time series (not the monthly climatology) as for the per pixel analysis. (Sect. 2.3, Supplement Part 1). The PHYSAT algorithm frequency of diatom detection variable (Table 1) is characterized by a large degree of sparsity and exact zeros that render the per-pixel DFT-based phenology analysis impossible. Regionally binned analysis addresses this sparsity issue. In order to link the NASW satellite data analysis to at-sea observations, Bermuda Atlantic Time Series (BATS) *in situ* primary production and sediment trap data were downloaded from http://bats.bios.edu. Details of BATS *in situ* methodology are given in the caption of Fig. S11.

Part 5. Details on Month of Maximum

Several algorithms that closely resemble Chl also exhibit minimal differences with the ensemble mean month of maximum (Fig. S2), especially BR10, OC-PFT, CB06, and to a lesser extent UITZ06 and MY10. This result is not surprising for the abundance-based algorithms (BR10, OC-PFT, UITZ06) (Table 1) as these are parameterized with Chl; thus their retrievals are a strong function of Chl. For spectral-based models (CB06, MY10), this suggests the temporal variations of derived spectral absorption properties of phytoplankton are generally consistent with those of Chl (i.e. the absorption spectrum becomes flatter, indicating, as generally expected, an increasing proportion of microplankton when Chl concentration increases). In contrast, the remaining algorithms exhibit larger excursions from the ensemble mean, with the biggest differences occurring in the Southern Ocean for KSM09 and ROY13 (both exhibiting a similar pattern), the subtropical gyres for FUJI11, and many widespread areas for ROY13 and PhytoDOAS. Differences are expected for FUJI11 in the gyres as this algorithm is developed with high-latitude regional data only. PhytoDOAS differences may be attributable to some extent to the use of data from a different sensor with a different spatial and temporal resolution, and retrieving diatom Chl and not microphytoplankton fraction. It is not clear how to explain the differences in ROY13. Overall, since secondary blooms are also detected (Sects. 2.3 and 3.5; this Supplement Part 7), it is possible that some secondary and primary blooms are hard to distinguish and can be confused due to data noise, if their amplitudes are similar. This is particularly true when dealing with fractional biomass, as opposed to absolute units (*Cabré et al., 2016*). This similarity in amplitude is the particularly evident in KSM09 data in the temperate

and subarctic North Atlantic (e.g. Fig. S1). Note that since KSM09 is based on backscattering, it may be detecting phases of the bloom differently, e.g. because peak carbon may not be the same as peak Chl.

Part 6. Details on Primary Bloom Duration

Maps of primary bloom duration for Chl and the individual PFT algorithms (Fig. S4) reveal significant differences among them, as well as significant variability of high spatial frequency (likely noise). The spatial patterns for Chl determined with the DFT analysis here closely agree with the observations of *Racault et al. (2012)* (see their Fig. 1d). The Chl spatial patterns of Fig. S4 (as well as the ensemble mean PFT–based patterns of Fig. 4A) also agree to first order with the results of *Sapiano et al. (2012)*. Namely, *Sapiano et al. (2012)* also observe zonal maxima at ~30^{o} N/S latitude as the most prominent feature of the bloom duration global map. Durations there are quantified at ~ 170-180 days (see their Fig. 9), roughly consistent with the long bloom durations (> 120 days, often up to 180 days) as quantified here by the DFT method for Chl (Fig. S4, top left panel). Other features are also in agreement, such as the long bloom durations equatorward of the subtropical gyres, in the entire South Atlantic gyre, as well as in parts of the subarctic Atlantic. As a quantification of the level of agreement in terms of bloom duration, the standard deviations of bloom duration for the satellite algorithms are depicted in Fig. S5A, and for the CMIP5 models – in Fig. S5B.

As with other phenology metrics, the abundance-based PFT algorithms are most similar to Chl (BR10 and OC-PFT in particular, and UITZ06 to a lesser extent). CB06 is also similar, indicating consistency between the spectral shape of absorption coefficients and Chl concentration. ROY13 and KSM09 exhibit long bloom durations in the Southern Ocean, unlike the rest of the algorithms. PhytoDOAS, MY10, and FUJI11 exhibit significantly shorter bloom durations overall globally, as compared to the other algorithms. Overall, there is considerable disagreement among the PFT algorithms with regards to the primary bloom duration metric. The standard deviation of bloom duration among the 10 PFT algorithms tends to be about 20 – 40 days over much of the ocean, but it can reach up to 70 days in certain areas (Fig. S5A). The 7 CMIP5 models also exhibit high standard deviation of primary bloom duration (Fig. S5B); however, most high values occur in the tropics and subtropics, whereas higher latitudes generally exhibit better model agreement with regards to bloom duration (but this observation could be influenced by missing data at high latitudes, Fig. S6).

Part 7. Details on Secondary Blooms

Note that the maps of Fig. 5A and Fig. 5B are not exactly complementary to each other (one is not equal to unity minus the other) as some areas exhibit a complex signal with more than two peaks which may or may not be ecologically significant (they are not analyzed here). The model of *Sapiano et al. (2012)* detects a double peak in SeaWiFS Chl data in the Pacific at 40^{o}N, which is stranded by a single peak zone to the south and a flat model (no peak) to the north. This is roughly consistent with the PFT-based analysis presented here, except that the flat zone is identified as a single peak zone, and the PFT data has high data sparsity further north. The subtropical gyre and the Equatorial region in the Pacific exhibit many pixels of double peaks according to most PFT algorithms in the North Pacific, unlike *Sapiano et al. (2012)*, who identify a flat seasonal cycle there. Note that *Sapiano et al. (2012)* choose among 8 different

models (including sinusoidal and flat models, and including a secular trend), which is a different methodology from the analysis employed here.

Additional quantitative description of the secondary bloom can be provided by the phase difference between it and the primary bloom, as well as the relative strength of the secondary bloom compared to the primary one. The phase difference between the primary and secondary bloom, i.e. their relative timing, is given here as the difference in months between their peak timing (Fig. S7). The mean fractional prominence of the secondary blooms across the 10 PFT algorithms (Fig. S8A) indicates that in most of the areas where secondary blooms tend to occur (Fig. 5B) fractional prominence tends to be between 30 % and 60%. The equatorial secondary blooms tend to be of high fractional prominence, suggesting two annual blooms of roughly equal strength. The CMIP5 model's mean fractional prominence has a very different spatial pattern (Fig. S8B), again emphasizing a lack of the secondary bloom at temperate latitudes in the models. In comparison to the mean of the 10 PFT algorithms, Chl phenology exhibits fewer places with a secondary bloom (Fig. S8C). Importantly, the North temperate Atlantic area that does have Chl secondary blooms, exhibits smaller fractional prominence than the corresponding satellite algorithm microplankton ensemble mean (cf. Fig. S8A). This is most likely due to the fact that fractional microplankton for most PFT algorithms is compared to absolute Chl units. There is evidence that secondary blooms tend to be more dominated by larger phytoplankton than the corresponding primary bloom (*Sommer, 1996; Cabré et al., 2016*), so the secondary blooms appear *more pronounced* (i.e. of higher relative prominence) if they are expressed in terms of microplankton fraction. Thus, phenology analysis may be different if absolute biomass or Chl is used in the analysis instead (*Cabré et al., 2016*).

Part 8. Details on Regionally Binned and BATS Time Series Analysis.

A regime equivalent to the Atlantic transitional subpolar-subtropical regime discussed in Sect. 3.6 exists in the North Pacific as well as in the Southern Ocean at 35-50^{o}S. Bimodal peaks result from light limitation in winter, growth in spring, then nutrient limitation in the summer and growth in the fall. The dynamics of the North Atlantic transitional and subpolar regimes is explained by *Evans and Parslow (1985)* and agrees with the Sverdrup critical depth theory (*Sverdrup, 1953*). Deep wintertime mixing ensures light limitation and little production and zooplankton population in winter, despite high nutrient supply. High nutrients from the winter and slow recovery of zooplankton in the spring and/or a large zooplankton class that does not respond fast enough to growing phytoplankton populations allow a spring phytoplankton bloom of large diatoms o*r Phaeocystis*. This bloom is terminated by a drop in nutrients and zooplankton grazing in the summer; the fall bloom starts when mixing re-introduces nutrients to the upper layer. The seasonally varying *Northern subtropics* regime centered around 30^{o}N, is characterized, just like in the Pacific, by a single annual peak in winter or early spring and high seasonal variability. The NASW province and the BATS station (Fig. S9) discussed below are part of this regime; the Pacific Ocean shows an equivalent regime. Previous work has determined that this peak is due to seasonal entrainment of nitrogen into the mixed layer.

Fig. S11 shows a complementary analysis of satellite and *in situ* time-series for the nutrient-limited, subtropical BATS station located inside the NASW region (Fig. S9). Most algorithms (and CMIP5 models) show a single biomass spring peak in Feb - March, as expected from *in situ* observations in this well-studied region (for a review see *Lomas et al. 2013*). The climatologies

of biomass and Chl are similar at BATS and NASW (compare Fig. 6B and Fig. S11A). As expected, the algorithms show slightly more consistency in the timing of the spring peak at BATS compared to the larger NASW region. Mixed layers at BATS vary from 10 m in the summer to 100 - 400 m in the winter, depending on the strength and phase of the North Atlantic Oscillation (*Bates, 2012*). The spring peak in biomass is driven by vertical mixing bringing macronutrients into the euphotic zone during winter. This nutrient flux supports a short spring-bloom period of higher primary production (Fig. S11C, bottom) and enhanced chlorophyll and carbon in most taxa present (e.g. *Goericke, 1998*). The ROY13, FUJI11 and PhytoDOAS are, as in the NASW region case, the most anomalous compared to the mid-February biomass peak date in the algorithm mean (Fig S11B).

Part 9. Details on Sources of Uncertainty

Some additional disadvantages of the DFT technique apart from the ones discussed in Sect. 3.7) include the difficulty in providing confidence intervals, and (if the Fast Fourier Transform (FFT) algorithm is used for DFT computation, as it is here), the necessity for data to be sampled at regular intervals (i.e. necessitating interpolation over missing data and padding missing data at the edges of time series with zeros). In the high latitudes, interpolating over many winter months with missing data may dampen the amplitude and create artificial waveforms that appear as spurious energy in various frequencies (*Cole et al., 2012*; *Sapiano et al. 2012*). This is, however, a problem intrinsic to satellite observations, not the DFT technique. Long time series are required to achieve good frequency resolution (*deBeurs and Henebry, 2010*). The DFT technique can only distinguish frequencies that are half the sampling frequencies, i.e. up to the Nyquist frequency. Note that limitations such as aliasing (discussed in Sect. 3.7) and the Nyquist frequency are fundamental theoretical limitations of information and signal theory, rather than specific drawbacks of the DFT technique.

Just like in the oligotrophic tropics and subtropics, there is evidence that nanoplankton may dominate the annual bloom in the Southern Ocean as well (*Uitz et al., 2006; Sadeghi et al. 2012*). In the Southern Ocean, CMIP5 models and satellite data exhibit diverging relationships of total biomass to fractional biomass in different size classes (*Cabré et al., 2016*). Large differences of the various phenological parameters among the PFT algorithms in the Southern Ocean indicate that satellite data there need to be treated with caution and algorithm parameterizations need to pay special attention to this critical region. For example, it has been shown that for the OC-PFT algorithm other parameterizations are necessary when the algorithm is used in the Southern Ocean (*Soppa et al. 2014*). Parameterizations are generally challenging due to the dearth of *in situ* data from this remote region, and due to the tendency to build globally applicable algorithms. The Southern Ocean is known to be an atypical region in terms of bio-optics (*Uitz et al., 2006*) where satellite-derived Chl is generally underestimated (*Kahru and Mitchell, 2010*). Phytoplankton community structure could be one of the factors explaining this bias in ocean color Chl estimates (*Sathyendranath et al., 2001*; *Jonhson et al. 2013; Ward et al., 2015*), which itself is used to estimate PFTs in some algorithms. In addition, considerable contributions to backscattering by coccoliths (e.g. *Balch et al., 2005*; *Balch et al, 2011*) and bubbles (*Zhang et al., 2002*; *Randolph et al., 2014*) are known to occur in the Southern Ocean. Furthermore, due to the low sun angles and the polar night, as well as to considerable cloudiness, satellite data in the Southern Ocean is sparse and many algorithms do not have valid retrievals (Fig. S6), biasing spatial and temporal means. Thus there are several reasons for being cautious

when using satellite OCR-derived products in the Southern Ocean. For more details on the specificities of OCR in high latitudes, see *IOCCG (2015)*. Note that coccolithophores are fairly global in distribution and anomalous calcite backscattering can affect some algorithm results elsewhere, because it can introduce errors in band-ratio-derived Chl (*Balch et al., 2005*), or cause violations in the assumptions of the KSM09-based PFT retrievals, for example. In terms of phenology (specifically month of maximum), their confounding effect is likely mitigated to some degree depending on the algorithm, because calcite concentrations tends to co-vary with Chl (*Hopkins et al., 2015)*. Also note that some of the algorithms specifically retrieve coccolithophores as a PFT (PHYSAT, PhytoDOAS), and thus take their effects into account, and NASA standard processing implements a high coccolithophore concentration flag, thus masking out some of their confounding signal.

Algorithm users need to keep in mind that algorithms (especially those based on empirical relationships) can only be as good as the data sets used to develop them; *in situ* data set uncertainty translates to algorithm failure or uncertainty. Development data sets are not necessarily representative of the ecosystem states geographically or temporally even within the regions they span. Finally, ensemble means may be biased toward algorithms based solely on Chl as input (3 of the 10 PFT algorithms are Chl-based).

Oceanic ecosystems are expected to exhibit biennial variability (*Platt et al., 2009*) and more complex interannual variability due to climate oscillations such as ENSO and PDO, as well as secular trends due to climate change. Thus next steps in the PFT phenology analysis need to involve longer time series and allow for temporal evolution of phenology, e.g. via wavelet analysis (or studying the fractional frequencies of the DFT, or using a sliding DFT), and allow for a secular trends in the model (*Weatherhead et al., 1998; Sapiano et al., 2012*). However, a longer time series is recommended for that. Such a time series needs to be self-consistent, which requires seamless merging of several successive satellite mission. Apart from phenology, study of long term trends using seamlessly merged satellite data sets (*Maritorena et al., 2010*) is important, focusing on a power analysis of trend estimation (*Gerrodette, 1987*), i.e. predicting the length of record required to distinguish trends from noise in the data (e.g. *Beaulieu et al., 2013*; *Kostadinov and Lookingbill, 2015*).

Additional Supplement References*

(*references appearing in the main text and also cited in the Supplement are not repeated here)

Allen, F. H., & Johnson, O. W. E. N. (1991). Automated conformational analysis from crystallographic data. 4. Statistical descriptors for a distribution of torsion angles. *Acta Crystallographica Section B: Structural Science*, 47(1), 62-67.

Aumont O, Bopp L (2006), Globalizing results from ocean in situ iron fertilization studies, *Global Biogeochemical Cycles* 20:15. doi:10.1029/2005gb002591.

Balch, W. M., Howard Gordon, B. C. Bowler, D. T. Drapeau, E. S. Booth. 2005. Calcium carbonate measurements in the surface global ocean based on MODIS Data. *Journal of Geophysical Research-Oceans*. 110, C07001 doi:10.1029/2004JC002560.

Balch WM, Drapeau DT, Bowler BC, Booth ES, Lyczskowski E, Alley D (2011) The contribution of coccolithophores to the optical and inorganic carbon budgets during the Southern Ocean Gas Experiment: New evidence in support of the "Great Calcite Belt" hypothesis. *Journal of Geophysical Research-Ocean*s Special Issue. VOL. 116, C00F06, doi:10.1029/2011JC006941.

Bates, N. R.: Multi-decadal uptake of carbon dioxide into subtropical mode water of the North Atlantic Ocean, *Biogeosciences*, 9, 2649-2659, doi:10.5194/bg-9-2649-2012, 2012.

Beaulieu, C., Henson, S. A., Sarmiento, Jorge L., Dunne, J. P., Doney, S. C., Rykaczewski, R. R., and Bopp, L. (2013), Factors challenging our ability to detect long-term trends in ocean chlorophyll, *Biogeosciences*, 10, 2711-2724, doi:10.5194/bg-10-2711-2013, 2013.

Berens, P. (2009), CircStat: a MATLAB toolbox for circular statistics, J. Statistical Software, 31(10), 1-21.
Cole, H., S. Henson, A. Martin, and A. Yool (2012), Mind the gap: The impact of missing data on the calculation of phytoplankton phenology metrics, *J. Geophys. Res.*, 117, C08030, doi:10.1029/2012JC008249.

Claus S., De Hauwere N., Vanhoorne B., Hernandez F., Mees J. (Flanders Marine Institute) (2013), Marineregions.org, Accessed at http://www.marineregions.org on 2013-05-01.

Cole, H., S. Henson, A. Martin, and A. Yool (2012), Mind the gap: The impact of missing data on the calculation of phytoplankton phenology metrics, *J. Geophys. Res.*, 117, C08030, doi:10.1029/2012JC008249.

Gerrodette, T. (1987), A power analysis for detecting trends, *Ecology*, 1364-1372.

Goericke, R., (1998), Response of phytoplankton community structure and taxon-specific growth rates to seasonally varying physical forcing in the Sargasso Sea off Bermuda, *Limnology and Oceanography*, 43, 921-935.

Gregg WW (2008), Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model, *Journal of Marine Systems* 69:205-225. doi:10.1016/j.jmarsys.2006.02.015

Heinzel, G., A. Rudiger and R. Schilling (2002), Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows, Max-Planck-Institut fur Gravitationsphysik, (Albert-Einstein-Institut), Teilinstitut Hannover.

Hopkins, J., S. A. Henson, S. C. Painter, T. Tyrrell, and A. J. Poulton (2015), Phenological characteristics of global coccolithophore blooms, *Global Biogeochem. Cycles*, 29, 239–253, doi:10.1002/2014GB004919.

IOCCG (2015). Ocean Colour Remote Sensing in Polar Seas. Babin, M., Arrigo, K., Bélanger, S. and Forget, M-H. (eds.), IOCCG Report Series, No. 16, International Ocean Colour Coordinating Group, Dartmouth, Canada.

Kahru, M., and B. G. Mitchell (2010), Blending of ocean colour algorithms applied to the Southern Ocean, *Remote Sensing Letters*, 1(2), 119-124

Kostadinov, T. S. and Gilb, R. (2014), Earth Orbit v2.1: a 3-D visualization and analysis model of Earth's orbit, Milankovitch cycles and insolation, Geosci. Model Dev., 7, 1051-1068, doi:10.5194/gmd-7-1051-2014.

Kostadinov, T. S. and T. R. Lookingbill (2015), Snow cover variability in a forest ecotone of the Oregon Cascades via MODIS Terra products, *Remote Sensing of Environment*, 164C, pp. 155-169, doi:10.1016/j.rse.2015.04.002.

Lomas, M. W., Bates, N. R., Johnson, R. J., Knap, A. H., Steinberg, D. K., & Carlson, C. A. (2013). Two decades and counting: 24-years of sustained open ocean biogeochemical measurements in the Sargasso Sea, *Deep Sea Research Part II: Topical Studies in Oceanography*, 93, 16-32.

Maritorena, S., d'Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010), Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, 114(8), 1791-1804.

MathWorks(2015), Fast Fourier Transform (FFT), URL: http://www.mathworks.com/help/matlab/math/fast-fourier-transform-fft.html; Last access: Dec. 14, 2015.

Moody, A., & Johnson, D. M. (2001). Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sensing of Environment*, 75(3), 305-323.

Moore JK, Doney SC, Lindsay K (2004), Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, *Global Biogeochemical Cycles,* 18. doi:10.1029/2004gb002220.

Moore JK, Doney SC, Lindsay K, Mahowald N, Michaels AF (2006) ,Nitrogen fixation amplifies the ocean biogeochemical response to decadal timescale variations in mineral dust deposition, *Tellus Series B-Chemical and Physical Meteorology*, 58:560-572. doi:10.1111/j.1600-0889.2006.00209.x.

Palmer JR, Totterdell IJ (2001) Production and export in a global ocean ecosystem model, *Deep-Sea Res Part I-Oceanogr Res Pap,* 48:1169-1198. doi:10.1016/s0967-0637(00)00080-7.

Randolph, K., H.M. Dierssen, M. Twardowski, A. Cifuentes Lorenzen, C.J. Zappa (2014), Optical measurements of small deeply-penetrating bubble populations generated by breaking waves in the Southern Ocean. *J. Geophys. Res. Oceans*, 119, doi:10.1002/2013JC009227.

Riahi K et al (2011) RCP 8.5—a scenario of comparatively high greenhouse gas emissions. *Clim Change*, 109:33–57. doi:10.1007/s10584-011-0149-y

Sathyendranath, S., Stuart, V., Cota, G., Maas, H., Platt, T. (2001), Remote sensing of phytoplankton pigments: a comparison of empirical and theoretical approaches, *International Journal of Remote Sensing,* 22, 249–273.
Sommer, U. (1996). Plankton ecology: The past two decades of progress, *Naturwissenschaften*, 1270 83, 293–301. doi:10.1007/s001140050291.

Soppa M. A., Hirata T., Silva B., Dinter T., Peeken I., Wiegmann S., Bracher A. (2014) Global retrieval of diatoms abundance based on phytoplankton pigments and satellite. Remote Sensing 6: 10089-10106.

Sverdrup, H. U. (1953). On Conditions for the Vernal Blooming of Phytoplankton, *J. Cons. int. Explor. Mer, 18 (3): 287-295*. doi:10.1093/icesjms/18.3.287.

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design, *Bulletin of the American Meteorological Society*, 93:485-498. doi:10.1175/bams-d-11-00094.1.

Ward B.A. (2015), Temperature-Correlated Changes in Phytoplankton Community Structure Are Restricted to Polar Waters. *PLoS ONE* 10(8): e0135581. doi:10.1371/journal.pone.0135581

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X. L., Choi, D. S., Cheang, W. K., Keller, T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick, J. E. (1998), Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res.-Atmos.*, 103, 17149–17161, doi:10.1029/98jd00995.

Wieser, M. E., Holden, N., Coplen, T. B., Böhlke, J. K., Berglund, M., Brand, W. A., De Bièvre, P., Gröning, M., Loss, R.D., Meija, J., Hirata, T., Prohaska, T., Schoenberg R., O'Connor, G., Walczyk, T., Yoneda, S. and Zhu, X. K. (2013), Atomic weights of the elements 2011 (IUPAC Technical Report). *Pure and Applied Chemistry*, 85(5), 1047-1078.

Zhang, X., Lewis, M., Lee, M., Johnson, B., & Korotaev, G. (2002). The volume scattering function of natural bubble populations. *Limnology and Oceanography*, *47*(5), 1273-1282.

Supplement Figures

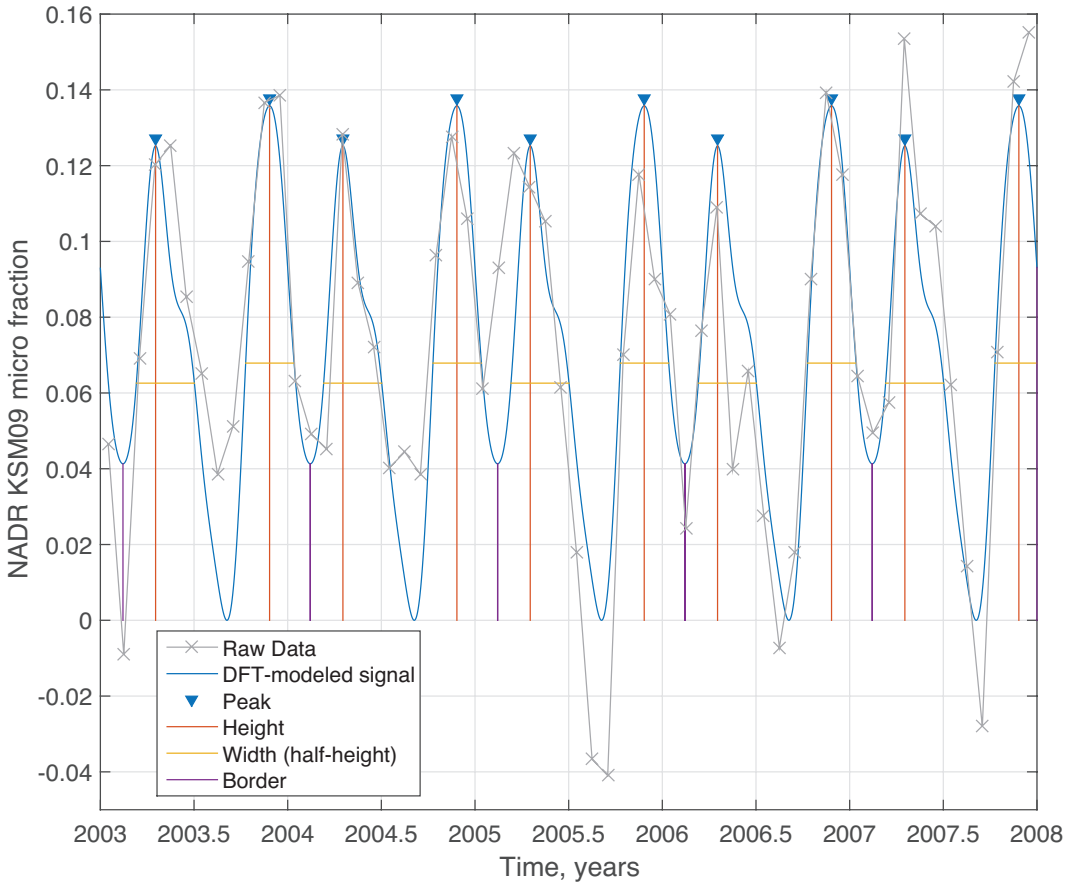

Figure S1. An example of DFT-based modeling of the annual seasonal cycle of PFT data, with subsequent peak analysis. The regionally binned KSM09 percent microplankton data for the *Longhurst (1998)* North Atlantic Drift province (NADR) is shown. The mean was subtracted from the data before modeling, and the minimum of the resulting modeled signal was subtracted before peak finding in order to identify height correctly. As a result, the modeled signal's minimum is zero, and some values of the raw data can be negative. Note the presence of two annual peaks of variable relative height in the data.

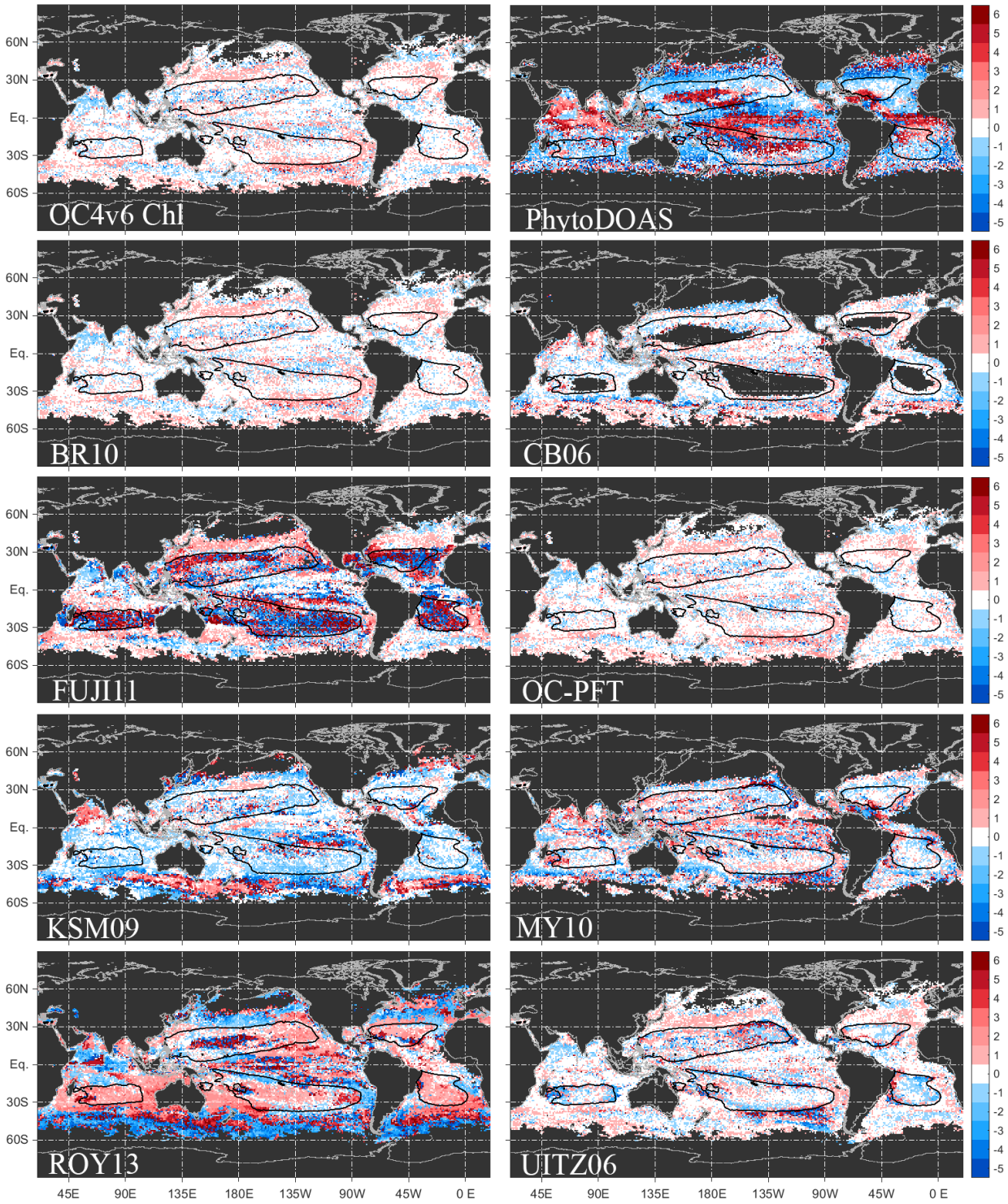

Figure S2. Maps of differences between the month of maximum bloom of OC4v6 Chl and the PFT algorithms (except PHYSAT, which exhibits very sparse data almost everywhere) and the ensemble mean of Fig 3A. Table 1 lists the algorithms and variables used. Positive differences indicate that the ensemble mean leads (i.e. occurs earlier than) the specific algorithm's month of maximum. Note that the determination of

month of maximum becomes unreliable when only a small percentage of the signal variance is explained by the seasonal cycle, thus these areas should be treated with caution (Figs. 2A, S13A). The isoline of climatological Chl = 0.08 mg m^{-3} is shown (black solid contour). Note that the CB06 algorithm retrieves almost no data in the subtropical gyres.

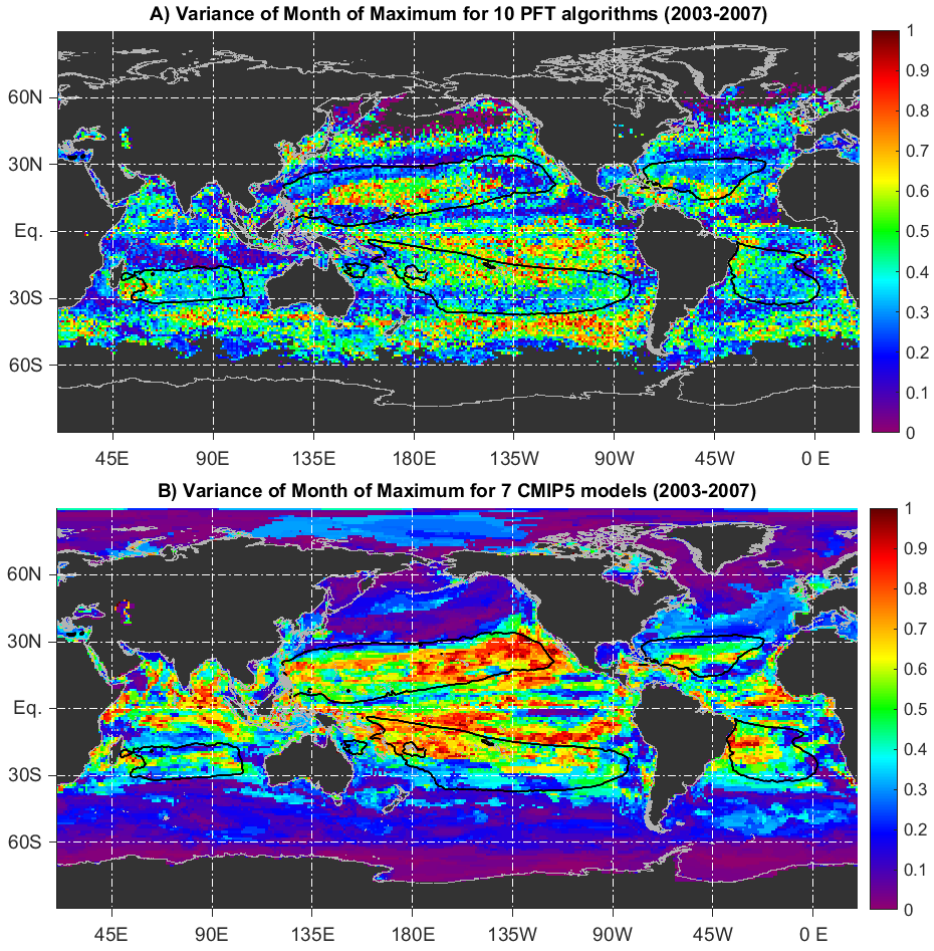

Figure S3. Circular variance of month of maximum of the primary bloom for (A) the 10 PFT algorithms and (B) the 7 CMIP5 models. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on both panels.

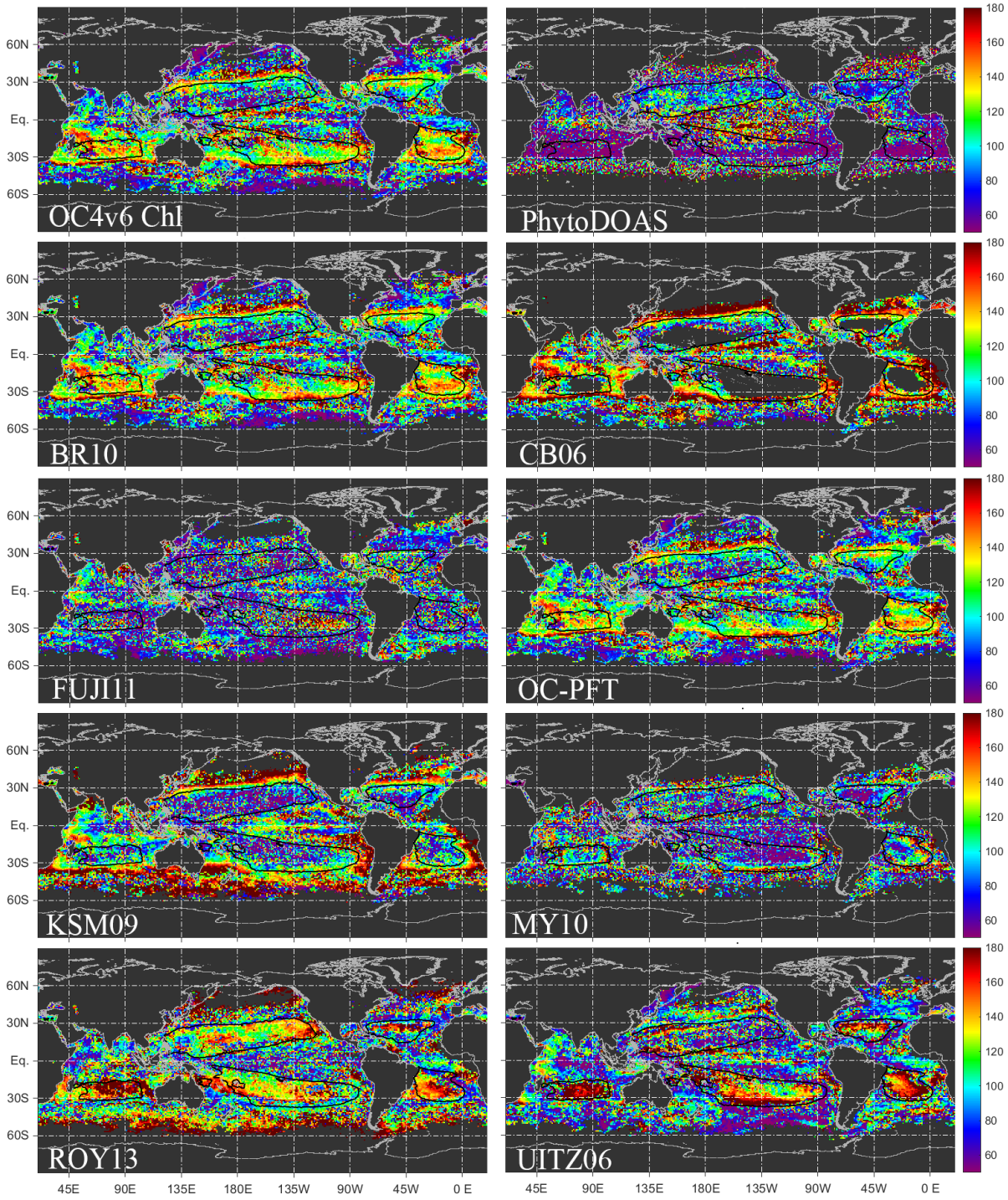

Figure S4. Maps of bloom duration (in days) for the primary bloom for OC4v6 Chl and the PFT algorithms (except PHYSAT). Duration of a bloom is defined as the width of the modeled seasonal signal at half the bloom peak height. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on all panels.

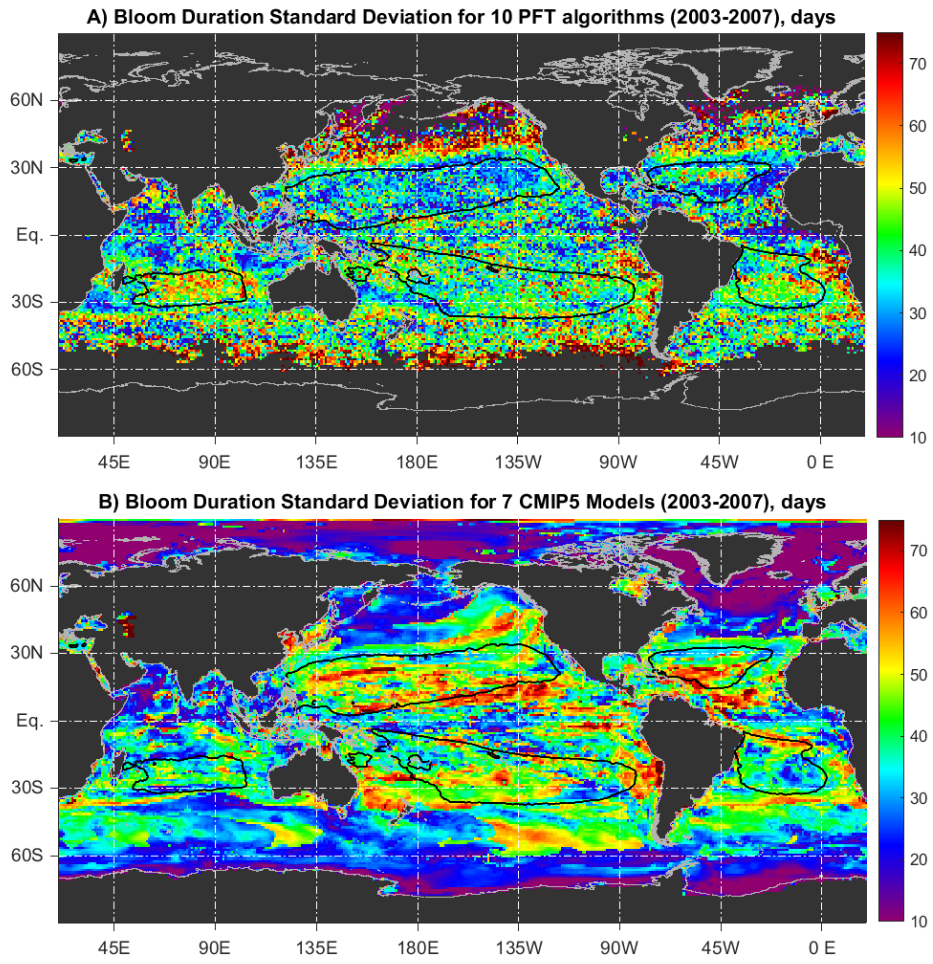

Figure S5. Standard deviation (in days) of the primary bloom duration for (A) the 10 PFT algorithms and (B) the 7 CMIP5 models. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on both panels.

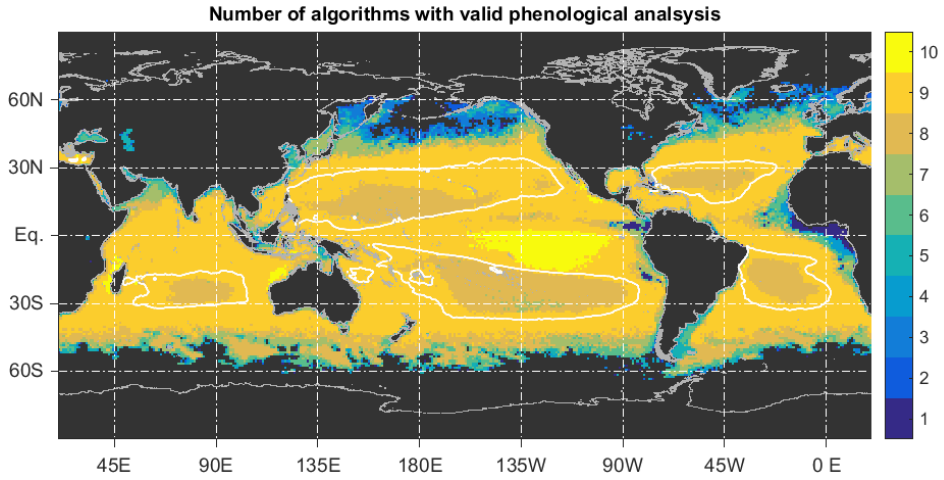

Figure S6. The number of algorithms for which valid phenological analysis is available. Compare with panels of Fig. 1, for example. The isoline of climatological Chl = 0.08 mg m^{-3} is shown as a white contour.

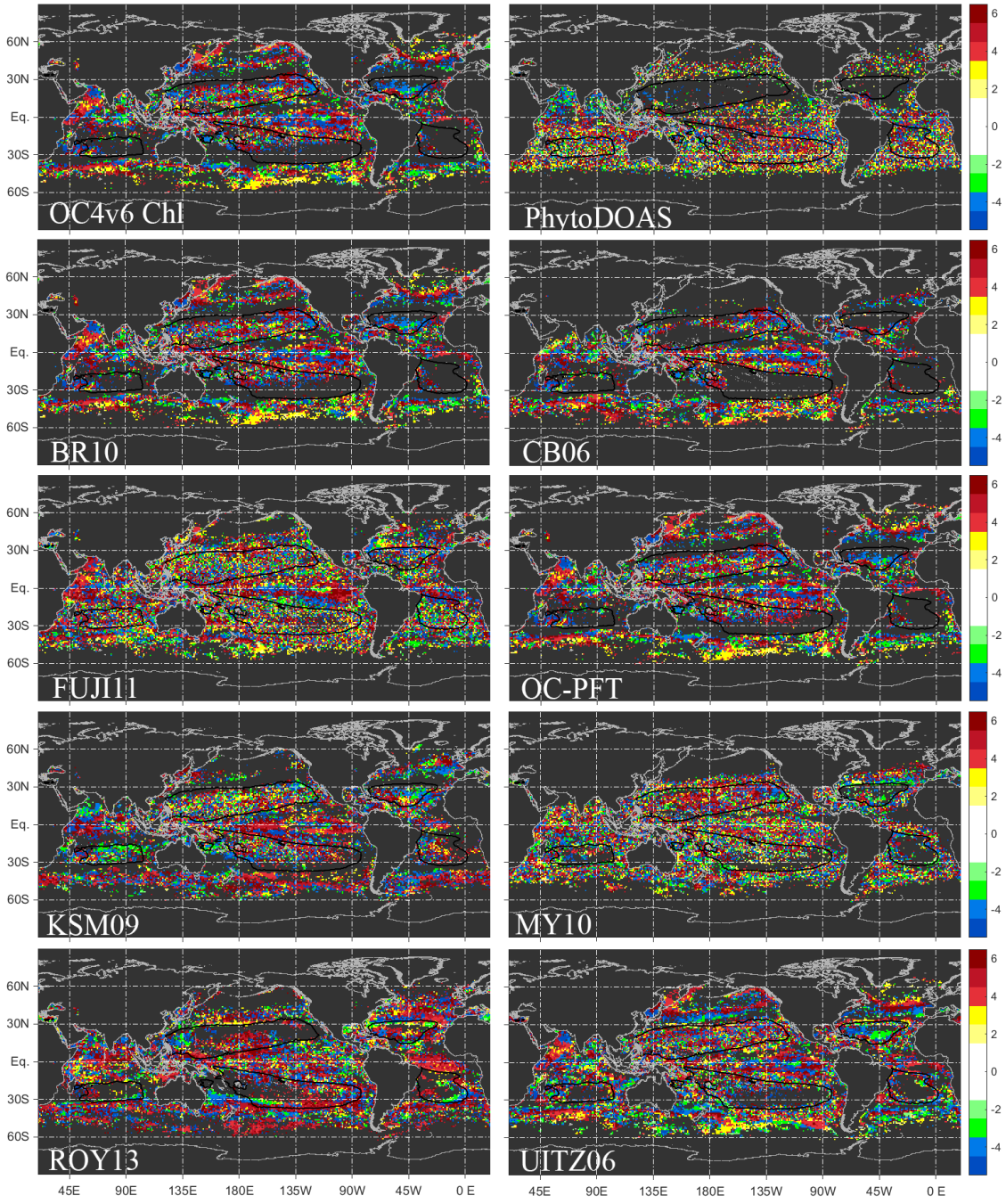

Figure S7. Difference in months between the maxima of the primary and the secondary blooms for OC4v6 Chl and the PFT algorithms (except PHYSAT). Differences greater than three months in absolute value are shown in either red colors for positive differences (the primary bloom leads, i.e. occurs earlier than the secondary bloom) or blue colors for negative differences (the secondary bloom occurs earlier than the primary bloom). A difference of six months is shown as positive by convention. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on all panels.

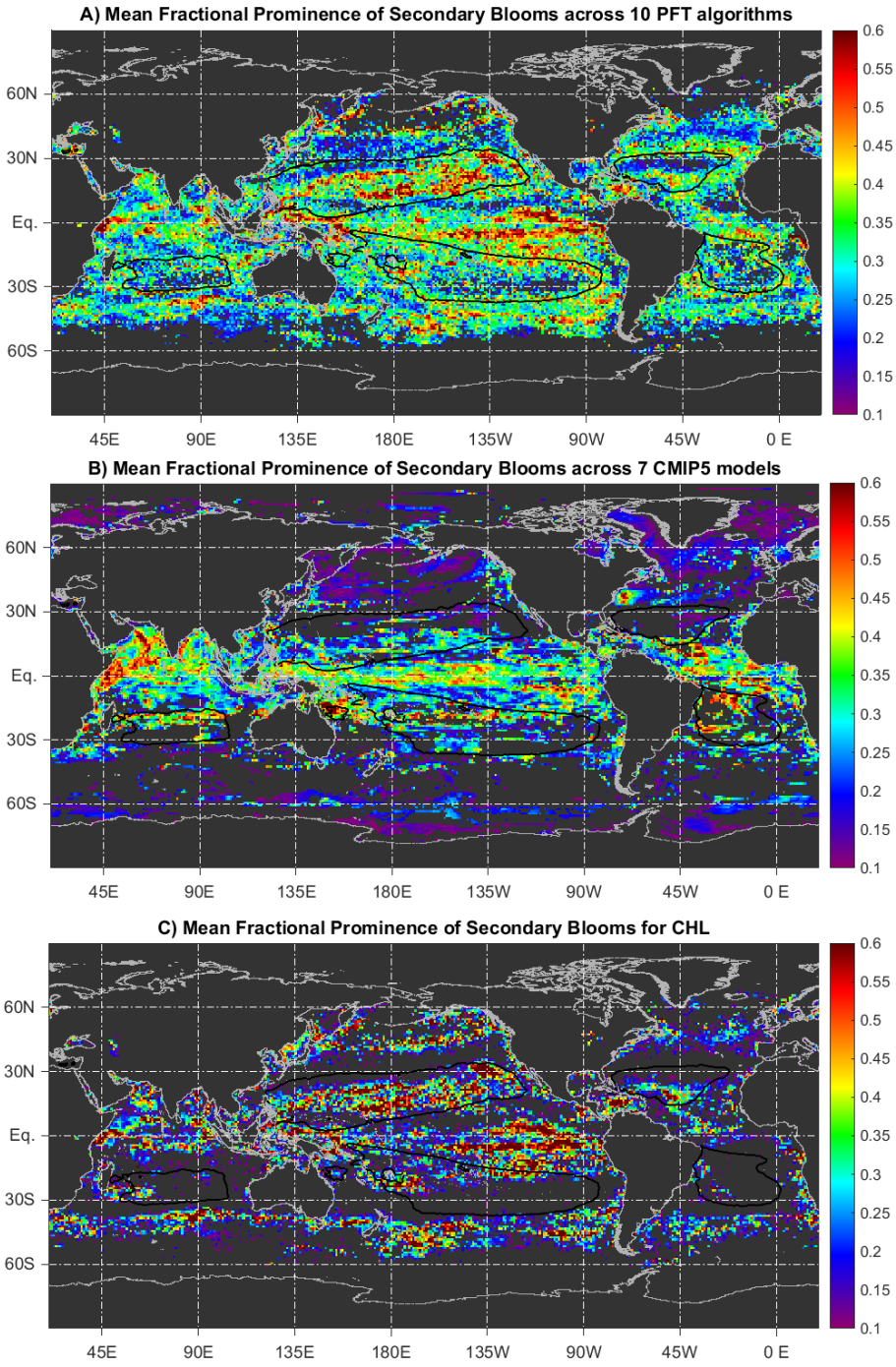

Figure S8. Fractional prominence of the secondary bloom in relation to the primary bloom, for, as follows: A) ensemble mean of percent microplankton for the 10 PFT algorithms, B) ensemble mean for diatom C for the 7 CMIP5 models, C) for OC4v6 Chl. Fractional prominence refers to the ratio of the prominence of the secondary bloom to the prominence of the primary bloom. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on all panels.

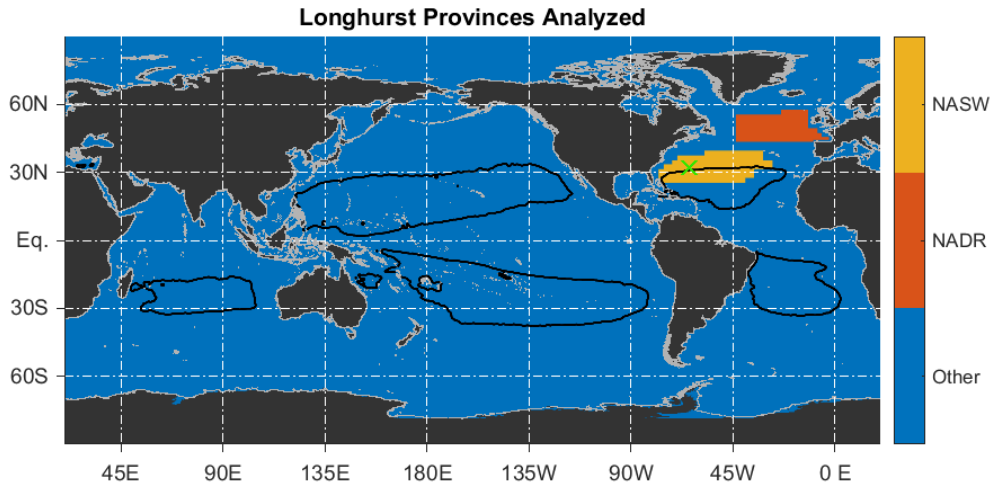

Figure S9. Map of the two *Longhurst (1998)* marine biogeographical provinces that were used for the regionally binned analysis – NADR (red) and NASW (orange). The location of the BATS station is indicated with a green cross (inside NASW). For details and province codes explanation, see this Supplement Part 4.

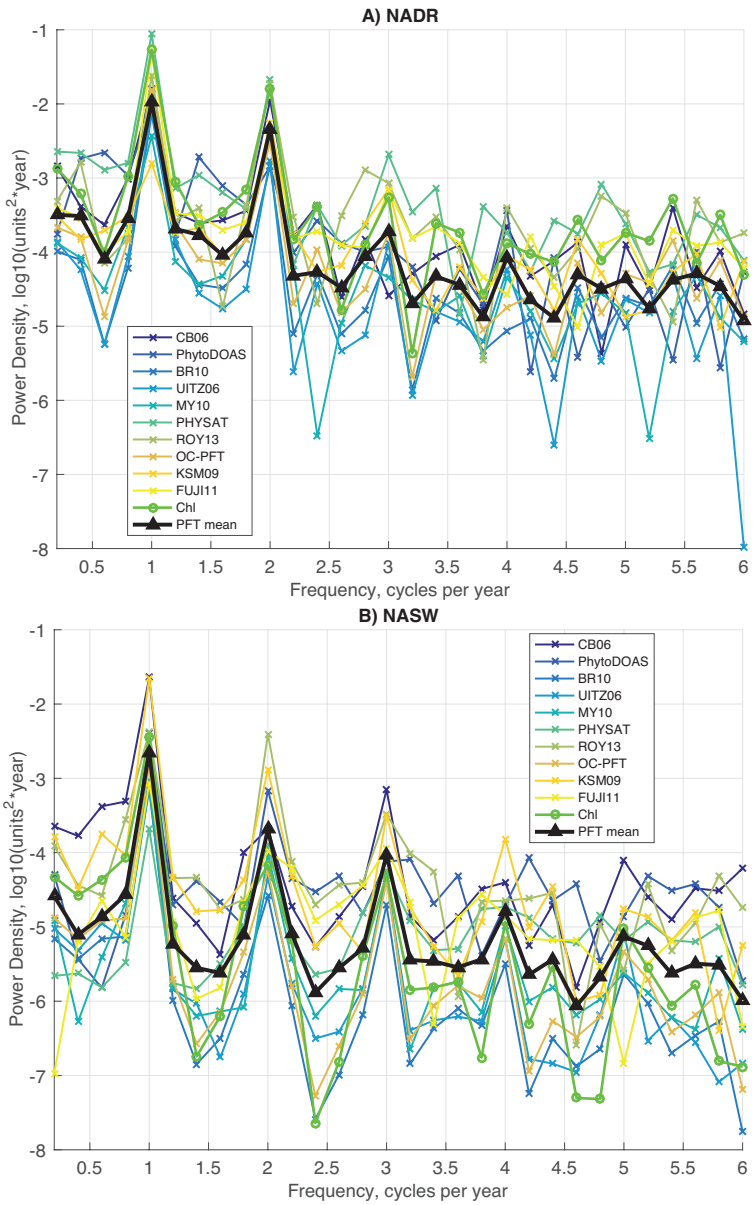

Figure S10. Power spectral density plots of the PFT algorithms (Table 1) and OC4v6 Chl for two example *Longhurst (1998)* biogeographic provinces as follows: A) North Atlantic Drift Region (NADR); B) The Western North Atlantic Subtropical Gyral Province (NASW). See Fig. S9 for a map of the provinces. See Sect. 2.3 and Supplement Part 1 for methodology details.

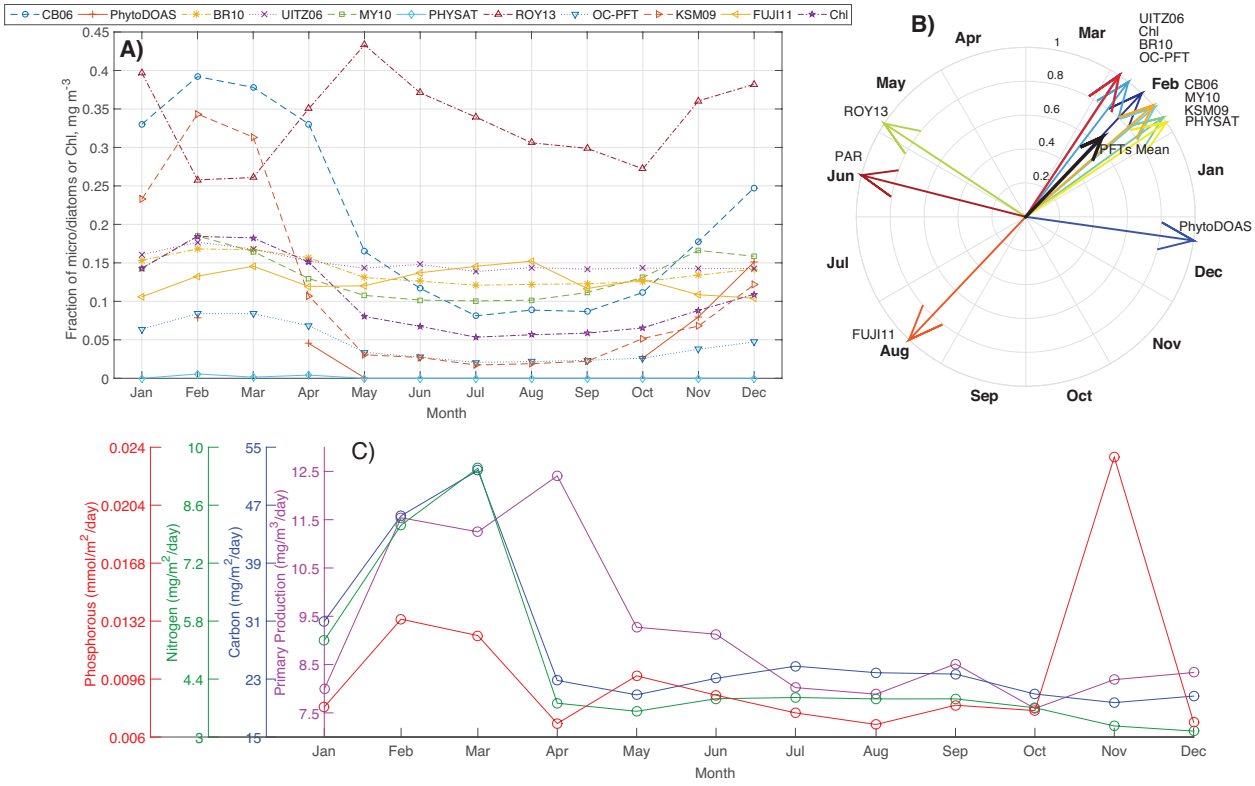

Figure S11. A) Time series as in Fig. 6, but for a 1x1° box centered on the BATS station only. B) As in Fig. 7, but for a 1x1° box centered on the BATS station only. C) Seasonal cycle of *in situ* biogeochemical data at the Bermuda Atlantic Time Series station at 31°50'N, 64°10'W. Purple: Primary Production, in mg m^{-3} day^{-1}. Go-Flo bottles were used to measure primary production at different depths. The shallowest depth for which there was consistent data, 5m, was used. Blue: Organic Carbon flux collected in sediment traps at the BATS station, in mg m^{-2} day^{-1}. Sediment traps were left underwater at specific depths for 3 days, usually around once per month. 3 traps were used, and the average was taken. Again, the shallowest depth for which there was consistent data, 150m, was used. Drop sites for a given day varied by as much as half a degree. The traps would also float up to a degree during their 3-day deployment. Green: Nitrogen flux, in mg m^{-2} day^{-1}, collected in the same traps as Carbon. Red: Phosphorous flux, in mmol m^{-2} day^{-1}, collected in the same traps as Carbon. Climatologies for each variable were computed over the complete SeaWiFS period (September 1997 to December 2010).

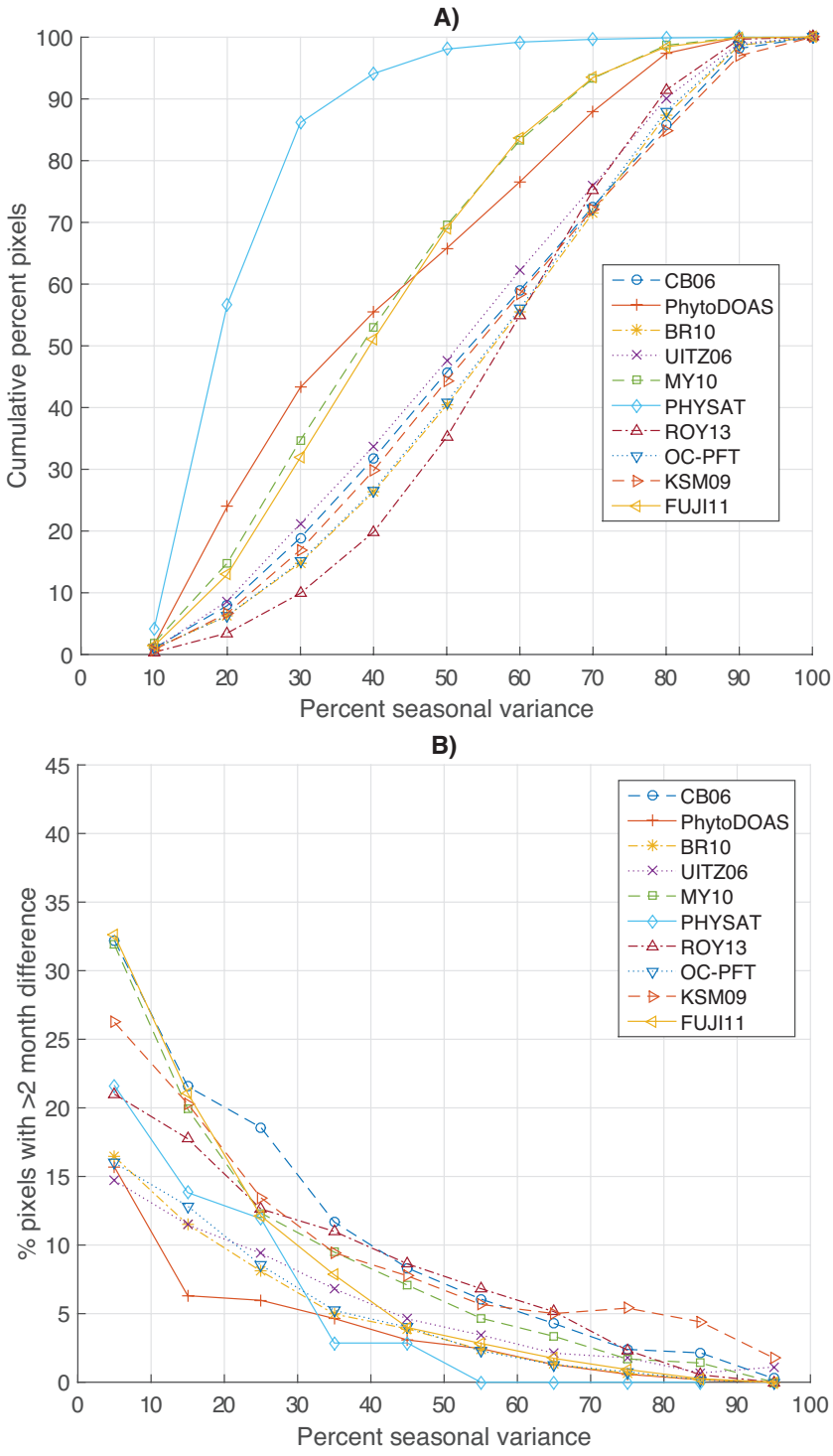

Figure S12. A) Cumulative percent of all pixels with valid phenological computations as a function of variance explained by the modeled seasonal cycle. For example, the graph indicates that for the MY10 model, ~70% of all analyzed pixels exhibit at most 50% variance explained by the modeled seasonal cycle. Note that differences among the algorithms can be intrinsic or can be due to differences in spatial coverage among the models (Fig. 1, Supplement Fig. S6), as illustrated prominently by the PHYSAT

curve. B) Percentage of pixels exhibiting greater than 2 months difference in month of annual maximum as determined from the DFT-based modeling used here, and as determined from direct peak analysis of the times series of the monthly climatologies of PFT algorithm satellite data (12 data points in each analyzed time series). The percentage is given as a function of percent variance explained by the DFT-modeled seasonal cycle, binned into 10%-bins. The percentage is calculated out of all pixels with valid phenology in each variance bin.

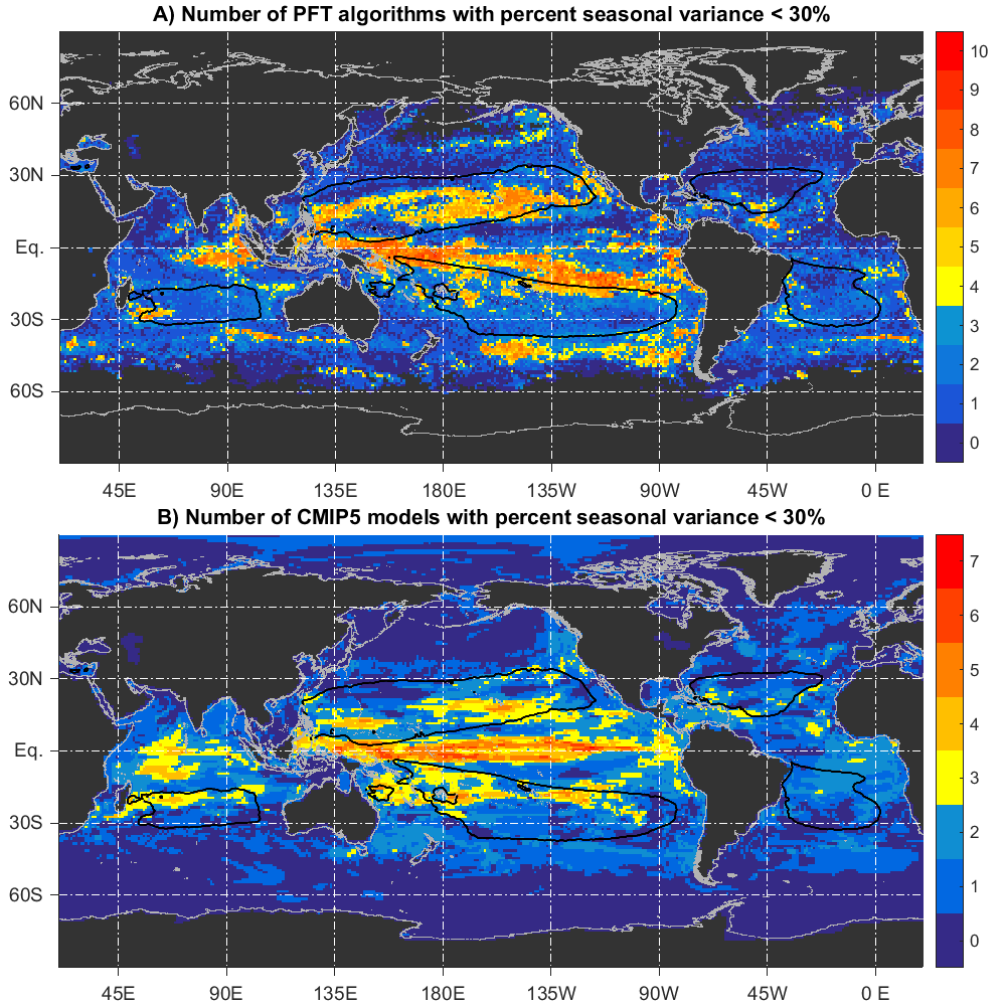

Figure S13. A) Number of PFT algorithms for which percent seasonal variance is less than 30%. Yellow and red colors indicate areas for which more than 3 PFT algorithms exhibit this, indicating that month of maxima (and other phenology metrics) should be interpreted with caution there. B) Same is an A), but for the 7 CMIP5 models. Yellow and red colors indicate areas where more than two models exhibit less than 30% seasonal variance. The isoline of climatological Chl = 0.08 mg m^{-3} (black solid contour) is shown on both panels.